WEEK 1

The Problem

Why Data Analysis?

- Data is everywhere.

- Data analysis/data science helps us answer questions from data.

- **Data analysis** plays an important role in:

  - Discovering useful information

  - Answering questions

  - Predicting future or the unknown

Tom wants to sell his car

Tom

How much money should he sell his car for?

The price he sets should not be too high, but not too low either.

Estimate used car prices

How can we help Tom determine the best price for his car?

- Is there data on the prices of other cars and their characteristics?

- What features of cars affect their prices?

  - Color? Brand? Horsepower? Something else?

- Asking the right questions in terms of data

Tom

Data on used car prices

```
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,13495
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,16500
1,?,alfa-romero,gas,std,two,hatchback,rwd,front,94.50,171.20,65.50,52.40,2823,ohcv,six,152,mpfi,2.68,3.47,9.00,154,5000,19,26,16500
2,164,audi,gas,std,four,sedan,fwd,front,99.80,176.60,66.20,54.30,2337,ohc,four,109,mpfi,3.19,3.40,10.00,102,5500,24,30,13950
2,164,audi,gas,std,four,sedan,4wd,front,99.40,176.60,66.40,54.30,2824,ohc,five,136,mpfi,3.19,3.40,8.00,115,5500,18,22,17450
2,?,audi,gas,std,two,sedan,fwd,front,99.80,177.30,66.30,53.10,2507,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,15250
1,158,audi,gas,std,four,sedan,fwd,front,105.80,192.70,71.40,55.70,2844,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,17710
1,?,audi,gas,std,four,wagon,fwd,front,105.80,192.70,71.40,55.70,2954,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,18920
1,158,audi,gas,turbo,four,sedan,fwd,front,105.80,192.70,71.40,55.90,3086,ohc,five,131,mpfi,3.13,3.40,8.30,140,5500,17,20,23875
0,?,audi,gas,turbo,two,hatchback,4wd,front,99.50,178.20,67.90,52.00,3053,ohc,five,131,mpfi,3.13,3.40,7.00,160,5500,16,22,?
2,192,bmw,gas,std,two,sedan,rwd,front,101.20,176.80,64.80,54.30,2395,ohc,four,108,mpfi,3.50,2.80,8.80,101,5800,23,29,16430
0,192,bmw,gas,std,four,sedan,rwd,front,101.20,176.80,64.80,54.30,2395,ohc,four,108,mpfi,3.50,2.80,8.80,101,5800,23,29,16925
0,188,bmw,gas,std,two,sedan,rwd,front,101.20,176.80,64.80,54.30,2710,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,21,28,20970
0,188,bmw,gas,std,four,sedan,rwd,front,101.20,176.80,64.80,54.30,2765,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,21,28,21105
1,?,bmw,gas,std,four,sedan,rwd,front,103.50,189.00,66.90,55.70,3055,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,20,25,24565
0,?,bmw,gas,std,four,sedan,rwd,front,103.50,189.00,66.90,55.70,3230,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,16,22,30760
0,?,bmw,gas,std,two,sedan,rwd,front,103.50,193.80,67.90,53.70,3380,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,16,22,41315
0,?,bmw,gas,std,four,sedan,rwd,front,110.00,197.00,70.90,56.30,3505,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,15,20,36880
2,121,chevrolet,gas,std,two,hatchback,fwd,front,88.40,141.10,60.30,53.20,1488,l,three,61,2bbl,2.91,3.03,9.50,48,5100,47,53,5151
```

https://archive.ics.uci.edu/ml/machine-learning-databases/autos/

Understanding the Data

Question

what does each line in the **csv** file represent

- ⦿ row
- ◯ column
- ◯ header

✓ **Correct**
correct, check the video

Skip   **Continue**

```
3,7,alfa-rom                              21,27,13495
3,7,alfa-rom                              21,27,16500
1,7,alfa-rom                              26,16500
2,164,audi,g
2,164,audi,g
2,7,audi,gas
1,158,audi,g
1,7,audi,gas
1,158,audi,g                              875
0,?,audi,gas
2,192,bmw,ga
0,192,bmw,ga
0,188,bmw,ga
0,188,bmw,ga
1,7,bmw,gas,
0,7,bmw,gas,
0,7,bmw,gas,
0,7,bmw,gas,
2,121,chevro                              5151
```

Each of the attributes in the dataset

No.	Attribute name	attribute range	No.	Attribute name	attribute range
1	symboling	-3, -2, -1, 0, 1, 2, 3.	14	curb-weight	continuous from 1488 to 4066.
2	normalized-losses	continuous from 65 to 256.	15	engine-type	dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
3	make	audi, bmw, etc.	16	num-of-cylinders	eight, five, four, six, three, twelve, two.
4	fuel-type	diesel, gas.	17	engine-size	continuous from 61 to 326.
5	aspiration	std, turbo.	18	fuel-system	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
6	num-of-doors	four, two.	19	bore	continuous from 2.54 to 3.94.
7	body-style	hardtop, wagon, etc.	20	stroke	continuous from 2.07 to 4.17.
8	drive-wheels	4wd, fwd, rwd.	21	compression-ratio	continuous from 7 to 23.
9	engine-location	front, rear.	22	horsepower	continuous from 48 to 288.
10	wheel-base	continuous from 86.6 120.9.	23	peak-rpm	continuous from 4150 to 6600.
11	length	continuous from 141.1 to 208.1.	24	city-mpg	continuous from 13 to 49.
12	width	continuous from 60.3 to 72.3.	25	highway-mpg	continuous from 16 to 54.
13	height	continuous from 47.8 to 59.8.	26	price	continuous from 5118 to 45400.

Attributes description: https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.names

Question

What is the name of the attribute what we want to predict

- ○ dataframe
- ○ feature
- ◉ target (label)

✓ **Correct**
correct

Skip   **Continue**

Each of the attributes in the dataset

No.	Attribute name	attribute range	No.	Attribute name	attribute range
1	symboling	-3, -2, -1, 0, 1, 2, 3.	14	curb-weight	continuous from 1488 to 4066.
2	normalized-losses	continuous from 65 to 256.	15	engine-type	dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
3	make	audi, bmw, etc.	16	num-of-cylinders	eight, five, four, six, three, twelve, two.
4	fuel-type	diesel, gas.	17	engine-size	continuous from 61 to 326.
5	aspiration	std, turbo.	18	fuel-system	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
6	num-of-doors	four, two.	19	bore	continuous from 2.54 to 3.94.
7	body-style	hardtop, wagon, etc.	20	stroke	continuous from 2.07 to 4.17.
8	drive-wheels	4wd, fwd, rwd.	21	compression-ratio	continuous from 7 to 23.
9	engine-location	front, rear.	22	horsepower	continuous from 48 to 288.
10	wheel-base	continuous from 86.6 120.9.	23	peak-rpm	continuous from 4150 to 6600.
11	length	continuous from 141.1 to 208.1.	24	city-mpg	continuous from 13 to 49.
12	width	continuous from 60.3 to 72.3.	25	highway-mpg	continuous from 16 to 54.
13	height	continuous from 47.8 to 59.8.	26	price	continuous from 5118 to 45400.

Target (Label)

Attributes description: https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.names

IBM Developer

SKILLS NETWORK

Coursera | Online Courses   ✕   Practice Quiz: Understandin ✕   +

coursera.org/learn/data-analysis-with-python/quiz/KYmrj/practice-quiz-understanding-the-data/attempt?redirectToCover=true

Python Tutor co...   C Program to Ch...   Other bookmarks

← **Back**

Practice Quiz: Understanding the Data
Practice Quiz • 6 min • 1 total point

✓ **Congratulations! You passed!**

Grade received 100%   To pass 60% or higher

Go to next item

1. Each column contains a:

◉ attribute or feature

◯ different used car

✓ **Correct**
correct

1 / 1 point

Scientifics Computing Libraries in Python

1. **Scientifics Computing** Libraries

Pandas
(Data structures & tools)

NumPy
(Arrays & matrices)

SciPy
(Integrals, solving differential equations, optimization)

Question

What is the primary instrument used in Pandas?

○ Arrays

◉ Dataframes

○ Matrices

✓ **Correct**
correct

Skip   **Continue**

Visualization Libraries in Python

2. **Visualization** Libraries

Matplotlib
(plots & graphs, most popular)
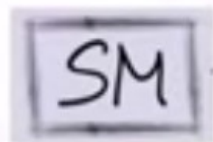
Seaborn
(plots : heat maps, time series, violin plots)

Algorithmic Libraries in Python

3. **Algorithmic** libraries

Scikit-learn
(Machine Learning : regression, classification,...)

SM

Statsmodels
(Explore data, estimate statistical models, and perform statistical tests.)

Question

what libraries do you use for data visualization.:

- ⦿ matplotlib
- ○ numpy
- ○ scikit-learn

✓ **Correct**
correct

Skip **Continue**

C Coursera | Online Courses   ✕   C Practice Quiz: Python Packa ✕   +

← → C ⌂   🔒 coursera.org/learn/data-analysis-with-python/quiz/99OuN/practice-quiz-python-packages-for-data-science/attempt?redirectToCover=true   ⟨ ☆   ✶ ◻ 🔵   Update ⋮

🐍 Python Tutor co...   ✂ C Program to Ch... | 🗁 Other bookmarks

Practice Quiz: Python Packages for Data Science

← **Back**   Practice Quiz • 2 min • 1 total point

✅ **Congratulations! You passed!**

Grade received **100%**   **To pass** 60% or higher

Go to next item

---

1. What is a Python library?

     1 / 1 point

⦿ A collection of functions and methods
      that allows you to perform lots of actions without writing your code.

◯ A file that contains data.

    ✓ **Correct**
      correct

Importing Data

- Process of loading and reading data into Python from various resources.

- **Two important properties:**

  - Format
    - various formats: .csv, .json, .xlsx, .hdf ….

  - File Path of dataset
    - Computer: /Desktop/mydata.csv
    - Internet: https://archive.ics.uci.edu/autos/imports-85.data

Getting Data

```
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,13495
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,16500
1,?,alfa-romero,gas,std,two,hatchback,rwd,front,94.50,171.20,65.50,52.40,2823,ohcv,six,152,mpfi,2.68,3.47,9.00,154,5000,19,26,16500
2,164,audi,gas,std,four,sedan,fwd,front,99.80,176.60,66.20,54.30,2337,ohc,four,109,mpfi,3.19,3.40,10.00,102,5500,24,30,13950
2,164,audi,gas,std,four,sedan,4wd,front,99.40,176.60,66.40,54.30,2824,ohc,five,136,mpfi,3.19,3.40,8.00,115,5500,18,22,17450
2,?,audi,gas,std,two,sedan,fwd,front,99.80,177.30,66.30,53.10,2507,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,15250
1,158,audi,gas,std,four,sedan,fwd,front,105.80,192.70,71.40,55.70,2844,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,17710
1,?,audi,gas,std,four,wagon,fwd,front,105.80,192.70,71.40,55.70,2954,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,18920
1,158,audi,gas,turbo,four,sedan,fwd,front,105.80,192.70,71.40,55.90,3086,ohc,five,131,mpfi,3.13,3.40,8.30,140,5500,17,20,23875
0,?,audi,gas,turbo,two,hatchback,4wd,front,99.50,178.20,67.90,52.00,3053,ohc,five,131,mpfi,3.13,3.40,7.00,160,5500,16,22,?
2,192,bmw,gas,std,two,sedan,rwd,front,101.20,176.80,64.80,54.30,2395,ohc,four,108,mpfi,3.50,2.80,8.80,101,5800,23,29,16430
0,192,bmw,gas,std,four,sedan,rwd,front,101.20,176.80,64.80,54.30,2395,ohc,four,108,mpfi,3.50,2.80,8.80,101,5800,23,29,16925
0,188,bmw,gas,std,two,sedan,rwd,front,101.20,176.80,64.80,54.30,2710,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,21,28,20970
0,188,bmw,gas,std,four,sedan,rwd,front,101.20,176.80,64.80,54.30,2765,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,21,28,21105
1,?,bmw,gas,std,four,sedan,rwd,front,103.50,189.00,66.90,55.70,3055,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,20,25,24565
0,?,bmw,gas,std,four,sedan,rwd,front,103.50,189.00,66.90,55.70,3230,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,16,22,30760
0,?,bmw,gas,std,two,sedan,rwd,front,103.50,193.80,67.90,53.70,3380,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,16,22,41315
0,?,bmw,gas,std,four,sedan,rwd,front,110.00,197.00,70.90,56.30,3505,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,15,20,36880
2,121,chevrolet,gas,std,two,hatchback,fwd,front,88.40,141.10,60.30,53.20,1488,1,three,61,2bbl,2.91,3.03,9.50,48,5100,47,53,5151
1,98,chevrolet,gas,std,two,hatchback,fwd,front,94.50,155.90,63.60,52.00,1874,ohc,four,90,2bbl,3.03,3.11,9.60,70,5400,38,43,6295
0,81,chevrolet,gas,std,four,sedan,fwd,front,94.50,158.80,63.60,52.00,1909,ohc,four,90,2bbl,3.03,3.11,9.60,70,5400,38,43,6575
1,118,dodge,gas,std,two,hatchback,fwd,front,93.70,157.30,63.80,50.80,1876,ohc,four,90,2bbl,2.97,3.23,9.41,68,5500,37,41,5572
1,118,dodge,gas,std,two,hatchback,fwd,front,93.70,157.30,63.80,50.80,1876,ohc,four,90,2bbl,2.97,3.23,9.40,68,5500,31,38,6377
1,118,dodge,gas,turbo,two,hatchback,fwd,front,93.70,157.30,63.80,50.80,2128,ohc,four,98,mpfi,3.03,3.39,7.60,102,5500,24,30,7957
1,148,dodge,gas,std,four,hatchback,fwd,front,93.70,157.30,63.80,50.60,1967,ohc,four,90,2bbl,2.97,3.23,9.40,68,5500,31,38,6229
1,148,dodge,gas,std,four,sedan,fwd,front,93.70,157.30,63.80,50.60,1989,ohc,four,90,2bbl,2.97,3.23,9.40,68,5500,31,38,6692
1,148,dodge,gas,std,four,sedan,fwd,front,93.70,157.30,63.80,50.60,1989,ohc,four,90,2bbl,2.97,3.23,9.40,68,5500,31,38,7609
1,148,dodge,gas,turbo,?,sedan,fwd,front,93.70,157.30,63.80,50.60,2191,ohc,four,98,mpfi,3.03,3.39,7.60,102,5500,24,30,8558
```

data source : https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data

Question

what is missing from the following code:

```python
1   import pandas as pd
2   # read the online file by the URL provided above, and assign it to variable "df"
3   path="https://archive.ics.uci.edu/ml/machine-learning-database/autos/imports-85.data"
4
5   df = read_csv(path,header=None)
```

○ read_csv should be pd.read_csv

○ there is no file path

○ pandas is imported as pd

✓ **Correct**
correct

Skip   **Continue**

Importing a CSV into Python

```python
import pandas as pd

url = "https://archive.ics.uci.edu/ml/machine-learningdatabases/autos/imports-85.data"

df = pd.read_csv(url)
```

Importing a CSV without a header

```python
import pandas as pd

url = "https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data"

df = pd.read_csv(url, header = None)
```

Question

what method was applied on the dataframe **df** to get the following output :

	0	1	2	3	4	5	6	7	8	9	...	16	17	18	19	20	21	22	23	24	25
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	13495
1	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	16500
2	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	154	5000	19	26	16500
3	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.40	10.0	102	5500	24	30	13950
4	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.40	8.0	115	5500	18	22	17450

- ⦿ df.head()

- ○ df.tail()

- ○ pd.head(df)

✓ Correct

Skip   Continue

Printing the dataframe in Python

- **df** prints the entire dataframe (not recommended for large datasets)

- **df.head(n)** to show the first *n* rows of data frame.

- **df.tail(n)** shows the bottom *n* rows of data frame.

`df.head()`

Header

	0	1	2	3	4	5	6	7	8	9	...	16	17	18	19	20	21	22	23	24	25
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	13495
1	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	16500
2	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	154	5000	19	26	16500
3	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.40	10.0	102	5500	24	30	13950
4	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.40	8.0	115	5500	18	22	17450

n=5

IBM Developer

SKILLS NETWORK

Adding headers

- Replace default header (by `df.columns = headers`)

```
headers = ["symboling","normalized-losses","make","fuel-type","aspiration","num-of-doors","body-style",
"drive-wheels","engine-location","wheel-base","length","width","height","curb-weight","engine-type",
"num-of-cylinders","engine-size","fuel-system","bore","stroke","compression-ratio","horsepower","peak-
rpm","city-mpg","highway-mpg","price"]
```

```
df.columns=headers
```

```
df.head(5)
```

Adding headers

- Replace default header (by `df.columns = headers`)

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore	stroke	compression-ratio	h
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	1
1	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	1
2	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	1
3	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.40	10.0	1
4	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.40	8.0	1

Exporting a Pandas dataframe to CSV

- Preserve progress anytime by saving modified dataset using

```
path="C:/Windows/…/ automobile.csv"
df.to_csv(path)
```

Exporting to different formats in Python

Data Format	Read	Save
csv	pd.read_csv()	df.to_csv()
json	pd.read_json()	df.to_json()
Excel	pd.read_excel()	df.to_excel()
sql	pd.read_sql()	df.to_sql()

C Coursera | Online Courses & ✕   C Practice Quiz: Importing and ✕   +

coursera.org/learn/data-analysis-with-python/quiz/IJM6y/practice-quiz-importing-and-exporting-data-in-python/attempt?redirectToCove...

Python Tutor co...   ⚙ C Program to Ch...   Other bookmarks

← **Back**

Practice Quiz: Importing and Exporting Data in Python
Practice Quiz • 3 min • 1 total point

✅ **Congratulations! You passed!**

Go to next item

Grade received 100%   **To pass** 60% or higher

1. What task does the following lines of code perform?

1 / 1 point

```
1    path='C:\Windows\…\ automobile.csv'
2    df.to_csv(path)
```

🔘 Exports your Pandas dataframe to a new csv file, in the location specified by the variable path.

⚪ Loads a csv file.

✅ **Correct**
   correct

Basic insights from the data

- Understand your data before you begin any analysis

- Should check:
  - Data Types
  - Data Distribution

- Locate potential issues with the data

Question

what data type do you expect the column that contains price to be

- ○ object
- ◉ float64
- ○ datetime64

✓ **Correct**
correct

Skip   **Continue**

Basic Insights of Dataset - Data Types

Pandas Type	Native Python Type	Description
object	string	numbers and strings
int64	int	Numeric characters
float64	float	Numeric characters with decimals
datetime64, timedelta[ns]	N/A (but see the datetime module in Python's standard library)	time data.

Why check data types?

- potential info and type mismatch

- compatibility with python methods

Basic Insights of Dataset - Data Types

- In pandas, we use `dataframe.dtypes` to check data types

`df.dtypes`

```
symboling            int64
normalized-losses    object
make                 object
fuel-type            object
aspiration           object
num-of-doors         object
body-style           object
drive-wheels         object
engine-location      object
wheel-base           float64
length               float64
width                float64
height               float64
curb-weight          int64
engine-type          object
num-of-cylinders     object
engine-size          int64
fuel-system          object
bore                 object
stroke               object
compression-ratio    float64
horsepower           object
peak-rpm             object
city-mpg             int64
highway-mpg          int64
price                object
dtype: object
```

dataframe.describe()

- Returns a statistical summary

```
df.describe()
```

	symboling	wheel-base	length	width	height	curb-weight	engine-size	compression-ratio	city-mpg	highway-mpg
count	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000
mean	0.834146	98.756585	174.049268	65.907805	53.724878	2555.565854	126.907317	10.142537	25.219512	30.751220
std	1.245307	6.021776	12.337289	2.145204	2.443522	520.680204	41.642693	3.972040	6.542142	6.886443
min	-2.000000	86.600000	141.100000	60.300000	47.800000	1488.000000	61.000000	7.000000	13.000000	16.000000
25%	0.000000	94.500000	166.300000	64.100000	52.000000	2145.000000	97.000000	8.600000	19.000000	25.000000
50%	1.000000	97.000000	173.200000	65.500000	54.100000	2414.000000	120.000000	9.000000	24.000000	30.000000
75%	2.000000	102.400000	183.100000	66.900000	55.500000	2935.000000	141.000000	9.400000	30.000000	34.000000
max	3.000000	120.900000	208.100000	72.300000	59.800000	4066.000000	326.000000	23.000000	49.000000	54.000000

Question

how would you generate descriptive statistics for **all** the columns for the dataframe **df**

○
```
1    df.describe()
```

◉
```
1    df.describe(include = "all")
```

○
```
1    df.info
```

Continue

dataframe.describe(include="all")

- Provides full summary statistics

`df.describe(include="all")`

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore	stroke
count	205.000000	205	205	205	205	205	205	205	205	205.000000	...	205.000000	205	205	205
unique	NaN	52	22	2	2	3	5	3	2	NaN	...	NaN	8	39	37
top	NaN	?	toyota	gas	std	four	sedan	fwd	front	NaN	...	NaN	mpfi	3.62	3.40
freq	NaN	41	32	185	168	114	96	120	202	NaN	...	NaN	94	23	20
mean	0.834146	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	98.756585	...	126.907317	NaN	NaN	NaN
std	1.245307	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6.021776	...	41.642693	NaN	NaN	NaN
min	-2.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	86.600000	...	61.000000	NaN	NaN	NaN
25%	0.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	94.500000	...	97.000000	NaN	NaN	NaN
50%	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	97.000000	...	120.000000	NaN	NaN	NaN
75%	2.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	102.400000	...	141.000000	NaN	NaN	NaN
max	3.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	120.900000	...	326.000000	NaN	NaN	NaN

Basic Insights of Dataset - Info

`dataframe.info()` provides a concise summary of your DataFrame.

```
df.info()
```

Row Number

C Coursera | Online Courses & X   C Practice Quiz: Getting Start X   +

coursera.org/learn/data-analysis-with-python/quiz/H6ZXs/practice-quiz-getting-started-analyzing-data-in-python/attempt?redirectToCov...

Python Tutor co...   C Program to Ch...   Other bookmarks

Practice Quiz: Getting Started Analyzing Data in Python

← **Back**

Practice Quiz • 3 min • 1 total point

✔ **Congratulations! You passed!**

Grade received 100%   **To pass** 60% or higher

Go to next item

1. To enable a summary of all the columns, what must the parameter **include** be set to for the method **describe?**

1 / 1 point

```
1    df.describe(include="all")
```

```
1    df.describe(include="None")
```
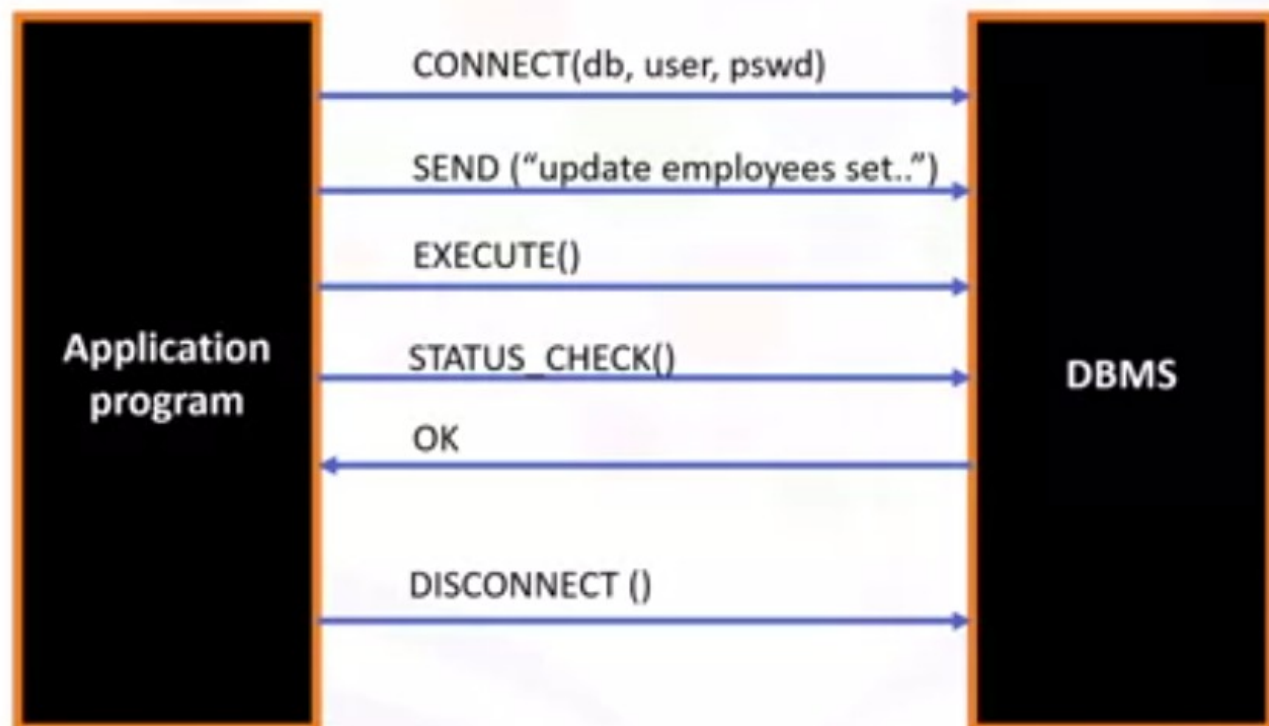
Accessing databases using Python

What is a SQL API?

What is a DB-API?

What is a DB-API?

Concepts of the Python DB API

Connection Objects

- Database connections
- Manage transactions

Cursor Objects

- Database Queries

What are Connection methods?

- cursor()
- commit()
- rollback()
- close()

Writing code using DB-API

```python
from dbmodule import connect

#Create connection object
connection = connect('databasename','username','pswd')

#Create a cursor object
cursor = connection.cursor()

#Run queries
cursor.execute('select * from mytable')
results = cursor.fetchall()

#Free resources
Cursor.close()
connection.close()
```

< Previous   Next >

Importing Datasets

✓ **Video:** The Problem
1 min

✓ **Video:** Understanding the Data
2 min

✓ **Practice Quiz:** Practice Quiz: Understanding the Data
1 question

✓ **Video:** Python Packages for Data Science
2 min

✓ **Practice Quiz:** Practice Quiz: Python Packages for Data Science
1 question

✓ **Video:** Importing and Exporting Data in Python
4 min

✓ **Practice Quiz:** Practice Quiz: Importing and Exporting Data in Python
1 question

Lesson Summary

In this lesson, you have learned how to:

Define the Business Problem: Look at the data and make some high-level decision on what kind of analysis should be done

Import and Export Data in Python: How to import data from multiple data sources using the Pandas library and how to export files into different formats.

Analyze Data in Python: How to do some introductory analysis in Python using functions like **dataframe.head()** to view the first few lines of the dataset, **dataframe.info()** to view the column names and data types.

Mark as completed